

University of Groningen

Observing, coaching and reflecting

van Helvert, Joy; Petukhova, Volha; Stevens, Christopher; de Weerd, Harmen; Börner, Dirk; van Rosmalen, Peter; Alexandersson, Jan; Taatgen, Niels

Published in:

EAI Endorsed Transactions on Future Intelligent Educational Environments

DOI:

[10.4108/eai.27-6-2016.151525](https://doi.org/10.4108/eai.27-6-2016.151525)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2016

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

van Helvert, J., Petukhova, V., Stevens, C., de Weerd, H., Börner, D., van Rosmalen, P., Alexandersson, J., & Taatgen, N. (2016). Observing, coaching and reflecting: Metalogue - A multi-modal tutoring system with metacognitive abilities. *EAI Endorsed Transactions on Future Intelligent Educational Environments*, 2016(6), [e6]. <https://doi.org/10.4108/eai.27-6-2016.151525>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Observing, Coaching and Reflecting: Metalogue - A Multi-modal Tutoring System with Metacognitive Abilities*

Joy Van Helvert^{1,*}, Volha Petukhova², Christopher Stevens³, Harmen de Weerd³, Dirk Börner⁴, Peter Van Rosmalen⁴, Jan Alexandersson⁵, Niels Taatgen³

¹University of Essex

²Spoken Language Systems, Saarland University, Germany

³Institute of Artificial Intelligence, University of Groningen

⁴Open University of the Netherlands

⁵German Research Centre for Artificial Intelligence GmbH

Abstract

The Metalogue project aims to develop a multi-modal, multi-party, dialogue system with metacognitive abilities that will advance our understanding of natural conversational human-machine interaction, and interfaces that incorporate multimodality into virtual and augmented reality environments. In this paper we describe the envisaged technical system, the learning contexts it is being developed to support and the pedagogical framework in which it is proposed user interactions will take place. This includes details of the system-generated learner feedback provided both in-performance and post-performance. We then move on to explain what has been achieved so far in terms of the integrated system pilots, and finally we discuss three key challenges the Metalogue researchers are currently working to overcome.

Received on 20 November 2015; accepted on 29 February 2016; published on 27 June 2016

Keywords: natural conversational interaction, mixed-reality, multi-modal dialogue systems, immersive, debate skills, learning analytics, reflection, negotiation.

Copyright © 2016 Joy van Helvert *et al.*, licensed to EAI. This is an open access article distributed under the terms of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi:10.4108/eai.27-6-2016.151525

1. Introduction

As we move towards a world of smart and immersive environments we are seeking new ways of interfacing and engaging with our technologies that more closely reflect natural human interaction. Human to human communication embodies multiple modalities such as speech, gesture, facial expressions, gaze, and body posture, so it follows there is an inherent desire to communicate with our technologies in the same way. This aspiration was illustrated recently in Spike Jonze' film "Her" [24] in which a writer encounters "Samantha"; a multimodal dialogue system capable of understanding, expressing, and responding to emotion to the extent that it was possible for him to fall in

love. While at present this level of immersion and engagement remains in the realms of science fiction, one of the aims of Metalogue, a three-year EU project, is to push the existing boundaries further in this direction.

So where to start? Absolutely free natural interaction is clearly not feasible at this point; however, educational dialogues and tutoring interventions offer some valuable constraints. The application of multimodal, natural language interfaces in this domain is already being explored, for example Nye et al. [32] text and speech, Johnson and Valente [23] discourse with "virtual humans", and Yang et al. [56] use of motion sensors; however, increasing computing power offers the possibility of combining a range of traditional and new modalities in a single dialogue system. In this paper we will outline the proposed Metalogue multimodal dialogue system and a potential application in the domain of metacognitive skills training. It has been shown that digital immersion can enhance learning in

*Please ensure that you use the most up to date class file, available from EAI at <http://doc.eai.eu/publications/transactions/latex/>

* Joy van Helvert. Email: jvanhe@essex.ac.uk

three ways: by providing (1) multiple perspectives; (2) situated learning; and (3) transfer [15, 27, 41]. While the vision for the Metalogue trainer may not be considered “digitally immersive” in the traditional sense, it can be said to offer an augmented reality that would incorporate these three types of experience.

We start by considering metacognitive skills training and how a multimodal dialogue system might identify and feedback salient issues in a learner’s performance. We look at two training settings—debate and negotiation interactions. This leads into an outline of the proposed physical system and details of its composition, and continues by to describe its implementation in the form of a metacognitive skills trainer discussing in detail, learning analytics and the design for reflective feedback. To close we review Metalogue’s immersive potential, report our progress so far in realising the vision and finally discuss three key issues Metalogue researchers are currently working on.

2. Metacognitive Skills Training

The added value of metacognitive skills and capabilities has long been researched in educational scenarios and dialogues, and recently, an increasing number of Intelligent Tutoring Systems (ITS), mostly developed by researchers in the U.S., have attempted to support metacognitive skills training [1, 4, 5, 10, 20]. Cognitive Tutor programs that represent innovative applications of technology, artificial intelligence and cognitive science have a sound theoretical basis for analysing how people learn [40]. However, to the best of our knowledge, none of these Cognitive Tutor Systems make use of natural spoken and multimodal dialogue. Research by Chi and colleagues [11] revealed that it is not so much what tutors do during tutoring sessions that is important, nor what students do, but how (and how much) the student and tutor respond to each other’s conversational moves. In other words, the interactivity of human tutoring drives its effectiveness. The interactivity of these tutoring systems, in our view, can be achieved through the use of multimodal, multimedia and multi-party dialogue, since such dialogue is the most social and natural form of interaction.

From the human perspective, the importance of metacognition in tutoring and learning has been proven in several psychological and pedagogical studies [4, 8, 9, 55]. Having better metacognitive skills can improve the student’s ability to learn and help them self-regulate learning across different domains and contexts [7]. As for the research evidence in favour of metacognition, results from a significant meta-analysis of instruction research conducted by Marzano [31] reported that approaches directed at the metacognitive level of setting goals, choosing appropriate strategies, and

monitoring progress are more effective in improving knowledge outcomes than those which simply aim to engage learners at the level of presenting information for understanding and use. Interventions targeted at improving metacognition produced an average gain of 26 percentile points (across 556 studies). This is about 5 points higher than the mean gain calculated for the 1772 studies in which attempts were made to improve cognition without an explicit metacognitive component. In this authoritative synthesis over 800 meta-analysis relating to achievement, Hattie [21] ranks meta-cognitive development programmes as the 13th most influential school innovation out of 138.

Metacognitive skills are domain independent and should be flexible enough to be applied while learning new domains and in a variety of different learning environments. Despite their transversal aspect, metacognitive skills learning should be situated and practised within certain domains and activity types. Regarding domains, there are systems already successfully used for mathematics, physics, geometry, and programming skills.

The Metalogue project focuses on two learning settings, debate and negotiation, that foster metacognitive skills and involve a bounded tutor providing learner interaction suitable for implementation within a multimodal, multimedia and multi-party dialogue system.

2.1. The Debate Setting

To debate successfully the student must master a range of metacognitive skills [29, 50], such as monitoring and adjusting verbal performance, eye contact, body posture and gestures, also they must know when and how to employ appropriate strategies and arguments to achieve certain goals while at the same time recognising and responding to the oppositions’ strategies and arguments. Debate skills training typically involves ad-hoc face-to-face classroom debates combined with more formal organised competitions. While the learner gains confidence through these performances, feedback from tutor or panel on such complex combined physical and mental skills played out over a period of time, depends on subjective human judgment and may reflect a summary of the interaction rather than pinpoint specific behaviours that can be improved.

The performance of a debater is often judged on three main criteria: (1) argument content; (2) argument organization and (3) argument delivery. Generally, the evaluation of argument content and its quality poses significant theoretical and computational challenges requiring a substantial amount of research and development. For instance, to detect logically flawed and/or irrelevant arguments inference machinery and consistency checking need to be implemented, and to

the best of our knowledge there is no system that is able to perform this task reliably.

As for organization of arguments, the planning and preparation of an Argument supported by Reason and Evidence are involved. Argumentative structures have been studied and modelled for argumentative texts and to a certain extent for two-party argumentative dialogues, see [33] for an overview. In order to identify arguments and relations between their constituents, discourse relations are often considered. Discourse relations help to identify to which other propositions a proposition serves as evidence and from which other propositions it receives support. For instance in [49], sentences within one argument and texts as a whole are classified as having one of the discourse relations such as result, purpose, background, solution, and scope achieving an F-score of 0.46.

Good debaters are distinguished by concise clear arguments connected by explicitly signalled structure, e.g. by discourse markers and dialogue act announcements. For example, “I will speak in favour of ... Because ... Since international research shows...”. For our task, we concentrate on detection of justification and evidence relations, and provide feedback to the learner whether the way she structured her arguments is in accordance with the tutoring system’s expectations.

Finally, in debate it is not only what arguments are brought up and how they are structured but also how they are presented or delivered that is important. In this respect, five aspects are considered: Audibility, Engagement, Conviction, Authority and Likability (AECAL). Good debaters should give a strong impression that they truly believe what they say. To express authority, confidence, respect and friendliness, the debater needs to use her body appropriately, control her voice, posture, emotions and maintain eye contact. More specifically, these important non-verbal communication aspects include gaze and facial expressions [3], head movements and orientation [17, 22], hand and arm gestures [25], and posture shifts or movements [42]. In the Metalogue project context the capturing and provisioning of feedback based upon these aspects has been outlined in [52] and further explored in [43].

2.2. The Negotiation Setting

As with successful debate, successful negotiation depends on the negotiator’s mastery of metacognitive skills. A successful negotiator is able to achieve an outcome that is beneficial to both the negotiator herself as well as her trading partner. The negotiator therefore has to balance competitive aspects of finding a solution that best suits her own goals, with cooperative aspects of finding mutually beneficial outcomes. It is important that the learner understands that a negotiation is not

a purely competitive interaction, and takes an active role in identifying the key goals of her trading partner to establish common ground. In addition, a good negotiator ensures that her trading partner perceives the negotiation as a positive experience as well.

The metacognitive skills that are involved in negotiation include aspects of (1) presentation, (2) interaction, and (3) content. Presentational aspects such as eye contact, tone of voice, body posture, and gestures help to determine a trading partner’s perception of the cooperativeness and trustworthiness of the negotiator. A highly cooperative negotiator can be judged as being stubborn or untrustworthy because of the way a learner presents herself. Through metacognitive training, a learner can become more aware of these aspects.

In addition, a learner’s perceived cooperativeness depends on the beliefs of her trading partner. Each negotiator typically has private information that is not available to her trading partner. A learner has to ensure that her trading partner obtains important information through proper interaction.

Finally, a good negotiator should pay attention to the content of her utterances. The end goal of a negotiator is to obtain as high a score for herself as possible. When none of the negotiators could have received a higher score without lowering the score of their trading partner, the outcome is said to be Pareto efficient. A good negotiator should be able to negotiate Pareto efficient outcomes. In addition, a good negotiator not only obtains a good outcome in the current negotiation, but also ensures that future negotiations with the same trading partner are possible. That is, a good negotiator is trustworthy in her negotiations. The trustworthiness of a negotiator can be judged by comparing her statements to her true valuations of different negotiation outcomes.

Practicing with a partner, even an artificial one, is an effective way to improve negotiation skills [28]. Metalogue goes a step further by providing a partner agent that actively reasons about and comments on the learner’s behavior. There is promising early evidence that metacognitive agents (i.e. those with the capacity to reason about the mind of the opponent) have an advantage in strategic games [48] and in negotiation [14]. Moreover, simply playing against such an agent can improve learners’ metacognitive reasoning in negotiation [13]. However, these existing agents do not provide explicit commentary or feedback on the learner’s behavior. The Metalogue system includes a metacognitive agent that is capable of providing feedback about the cooperativeness/aggressiveness of the learner’s strategy, its beliefs about the learner’s negotiation preferences, and the extent to which an optimal (pareto efficient) agreement was reached. Currently, the agent exists as a standalone application with a GUI interface [37], but in future versions of the

system the agent will be fully integrated with the rest of the system, allowing it to interpret the speech of the learner and respond using the system's avatar.

3. Metalogue: A Multimodal Dialogue System with Metacognitive Capabilities

The aim of Metalogue is to detect ("hear" and "see") a wide range of human interaction signals including those described above, and to "interpret", and "respond" in as natural way as possible, either in a coaching capacity or as a fully-fledged partner in the interaction. Here we describe the vision; the proposed systems architecture and workflows as depicted in Figure 1. The extent to which the vision has so far been realised is discussed in detail in Section 7.

The Metalogue dialogue system has two main tasks: (1) to track and understand human interaction behaviour in debate and/or negotiation scenarios, particularly focussing on the metacognitive or simple strategy elements; and (2) to perform certain tutoring interventions. 3.1. Tracking and interpretation Metalogue gathers three types of sensor specific data [44] from its input devices (see Figure 1. INPUT DEVICE layer): (1) speech signals from multiple sources, such as wearable microphones and headsets for each dialogue participant and, in the debate setting, an all-around microphone placed between participants; (2) visible movements in the form of tracking signals from Microsoft Kinect and Myo sensors capturing body movements and facial expressions; and (3) video signals captured by the camera and recording the whole dialogue including sound.

The speech signals serve as input for two types of further processing (see RECOGNITION LAYER and INTERPRETATION LAYER): Automatic Speech Recognition (ASR), leading to lexical, syntactic and semantic analysis to answer the question "what was said?"; and prosodic analysis which is concerned with rhythm, stress and intonation of speech and answers the question "how was it said?" This enables the system to interpret elements such as speech rate (fast, slow, adequate tempo), volume (loud, soft, adequate), emphasis (flat intonation, uneven/unbalanced intonation etc.), and pausing (too long or not enough). Analysis of visible movements gathered by Microsoft Kinect and Myo sensors, enable the system to interpret input related to gaze (re)direction, head movement and head orientation, facial expressions, hand and arm gestures, posture shifts, and body orientation. These outputs are further analysed to determine factors such as emotional state and, ultimately, argument content, organisation and delivery. By a combination of Fusion and Discourse Model, the modality interpretations are merged into

one homogeneous intention and given a context-dependent interpretation [38].

The combination of social and linguistic sign information is based on a large-scale human-human interaction data collection and annotation of data. This, in turn, is used to train machine-learning algorithms for the automatic recognition and prediction of a wide range of human interaction phenomena. For instance, the identification of arguments and analysis of their internal structure (i.e. evidence and justification relations from premises to conclusions) can be based on the identification and classification of discourse units and relations, and are learned in a data-oriented way, see [36]. Similarly, for the recognition of the intentions encoded in participants' utterances various machine learning techniques have been applied successfully such as Support Vector Machine (SVM) [6], Logistic Regression [57], AdaBoost [58] and the Linear Support Vector Classifier (LinearSVC)[53]. The F-scores ranging between 0.83 and 0.86 have been achieved which corresponds to the state-of-the-art performance [2].

Following [54], the system includes a Fusion and Generation Management component that given the systems intention(s), generate and split the intention(s) onto different modalities, such as Avatar and Voice (TTS), In-Action feedback (see Section 4.1.) and so on. The About-Action Feedback is generated from the content of the discourse model (see Section 4.1. and 4.2.) The Metalogue system makes use of the Display-as-a-Service (DaaS) technology in order for the visual output - Avatar, About-Action Feedback, to be flexibly rendered onto multiple screens.

3.1. Modelling, managing and generating tutoring interventions

Given the system's understanding of the learner's behaviour, its second task is to perform tutoring interventions in order to inform the learner of a mistake or to propose corrections or to provide positive feedback. Therefore, at the heart of the system (see CORE LAYER), interpretation and the generation of output depend on advanced linguistic multimodal and multi-perspective discourse models and dialogue management. Since in Metalogue we deal with several different tutoring aspects, an articulate discourse/dialogue model should contain all information considered relevant for interpreting such rich multimodal dialogue behaviour in order to enable the tutoring system to generate an adequate reaction. Tutoring interventions are then concerned with learner performance on argument and overall interaction structuring, fluency of spoken contributions, turn and time management, and managing perceptual and physical presentational aspects. In Metalogue, we designed a flexible model to deal

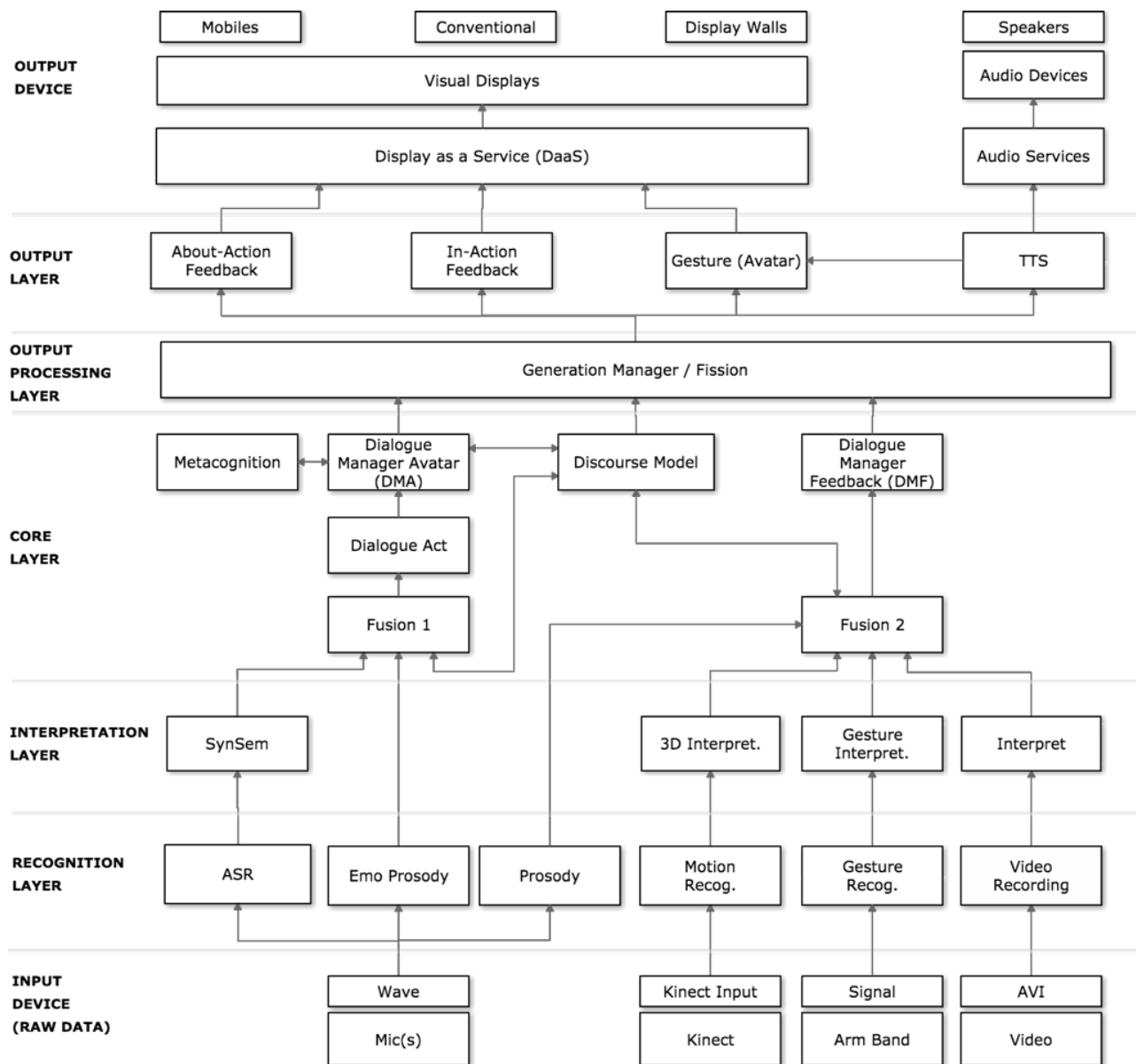


Figure 1. The technical architecture of the Metalogue system. From bottom to top, signals are reconized via input devices (“INPUT DEVICE” layer), further recognized by tailored analysis modules (“RECOGNITION” layer). After interpretation (“INTERPRETATION” layer), semantic representation from different modalities are fused (module “FUSION 1 and 2”) and given a context-dependent interpretation (module “DISCOURSE MODULE”) which eventually causes the “Dialogue Manager” and “Metacognition” modules to agree on actions. These are generated and fissioned onto different modalities (“OUTPUT PROCESSING” and “OUTPUT” layer) and finally rendered on the available output devices (“OUTPUT DEVICE” layer).

with such complex multimodal and multi-task communicative scenario. The proposed model has five components concerned with the actual dialogue (either debate or negotiation task, communicative (linguistic and/or multimodal) participants’ actions, perceptual information, participants’ cognitive processing states and social conventions such as rights and obligations in dialogue. Each of the parts of the model can be updated independently while other parts remain unaffected. For

instance, the linguistic part is updated when dealing with presentational aspects and some interaction aspects, such as turn management; in the cognitive part participants’ processing states are modelled, as well as aspects related to time and own communication management (e.g. error in speech production). The semantic part contains representations of task-related actions, in our scenario participants’ arguments and their structures, negotiation moves, and system’s tutoring goals and expectations of the learner’s learning progress.

The Metalogue Dialogue Manager, designed as a set of processes (threads), takes care of updating these various types of information and generates system's tutoring interventions, see [30] and [37]. The Metalogue approach supports more accurate understanding and multimodal and multi-tasking behaviour which is tuned to the debate/negotiation situation. Thanks to its domain and task-independent nature, the model offers possibilities for sophisticated refinements and structured extensions, but also for specific constraints, if required.

The unique contribution of the Metalogue system is the incorporation of metacognitive models that explain metacognition as a set of skills, a cognitive agent that exhibits metacognitive behavior similar to humans, and a learner model able to assess the users' metacognitive skills. Together these aspects of Metalogue enable the system to critically analyse participants' interactions within a certain time frame and generate "events" (i.e. points highlighting areas where performance could be improved and positive interaction behaviors that can be built upon). These are recorded in the form of annotations to the video file.

Output to the user can take three forms; (1) in-performance coaching; (2) post-performance reflective analysis; and (3) an optional debate or negotiation partner in the form of a virtual character that simulates a wide range of both verbal and non-verbal language attributes.

4. The Learning Context

Having outlined the technical design and the models that are key to the operation of Metalogue, we move on to consider how the learner and tutor might interact with one another and the system in a learning situation. Three context agnostic scenarios are envisaged; (1) tutor supported, (2) partially tutor supported, and (3) learner driven. All are set within the parameters of the 4C-ID pedagogical framework [51] which focuses on whole tasks that become more complex and are tackled with less support as competence increases. A task can be seen as a single round of interaction with the Metalogue system. We discuss here the tutor supported scenario illustrated in Figure 2, where the tutor plays an active role in agreeing objectives and setting up the path the learner will take (scenarios 2 and 3 are sub-sets of the tutor driven scenario that incrementally reduce tutor support in line with the learner's increasing competence). The boxes above the central line are the learner activities and those below are the tutor activities; boxes spanning the line are joint activities.

Moving from left to right, we start with the pre-assessment stage involving the human tutor and learner discussing the high level learning objectives. This

is followed by a Metalogue pre-assessment task in the form of a simple negotiation game that provides feedback on the learner's negotiation competence, and interaction with a presentation trainer to assesses the learners' habitual body posture and gestural skills. Metalogue will aim to provide a number of options that support tutor and learner in defining a virtual task or experience customised to the individual learners' needs. This includes defining the type of Metalogue feedback and/or tutoring response in terms of whether it will focus on developing the learners' presentational skills, interactional skills or content oriented skills. The system can also perform three different roles: (1) "Observer" (see Figure 3, left), in which the system is passive, recording and analysing interaction between two humans with the option to provide system generated in-action feedback (in-action and about-action feedback are discussed in detail in Section 5.); (2) "Experiencer" (see Figure 3, middle), where the system actively plays the role of one of the dialogue partners represented by an avatar; and (3); "Tutor" (see Figure 3, right), where the agent acts in a teaching capacity with the ability to guide multiple users including demonstrating good practice.

While Metalogue feedback is always individual to the learner, the ability to work with the tutor to define learning needs and then tailor the nature of the task accordingly enhances the relevance of the learning experience and increases the learner's sense of personal agency [12, 16]. In this scenario, the tutor sets up the dimensions of the task in the Metalogue system and the learner proceeds with the task in a single round of interaction.

As the learner performs the task, it is envisaged that the tutor will be able to annotate the recorded data via an iPad with her/his own simple key word observations. These will be visible as part of the post performance (About-action) feedback.

Finally, tutor and learner review the performance together supported by the comprehensive moment by moment analysis produced by the system in the form of a learning analytics dashboard (About-action feedback). Further rounds of interaction can be played with the same settings to gauge progress or new challenges can be created.

5. In-action and About-action Feedback

A key feature of Metalogue is its ability to provide two different forms of feedback designed to enhance the learner's experience [52] based on Donald Schön's [45] distinction between reflection in-action and about-action. Schön defines reflective practice as the practice by which professionals become aware of their implicit knowledge base and learn from their experience. He uses the terms reflection-in-action (reflection on

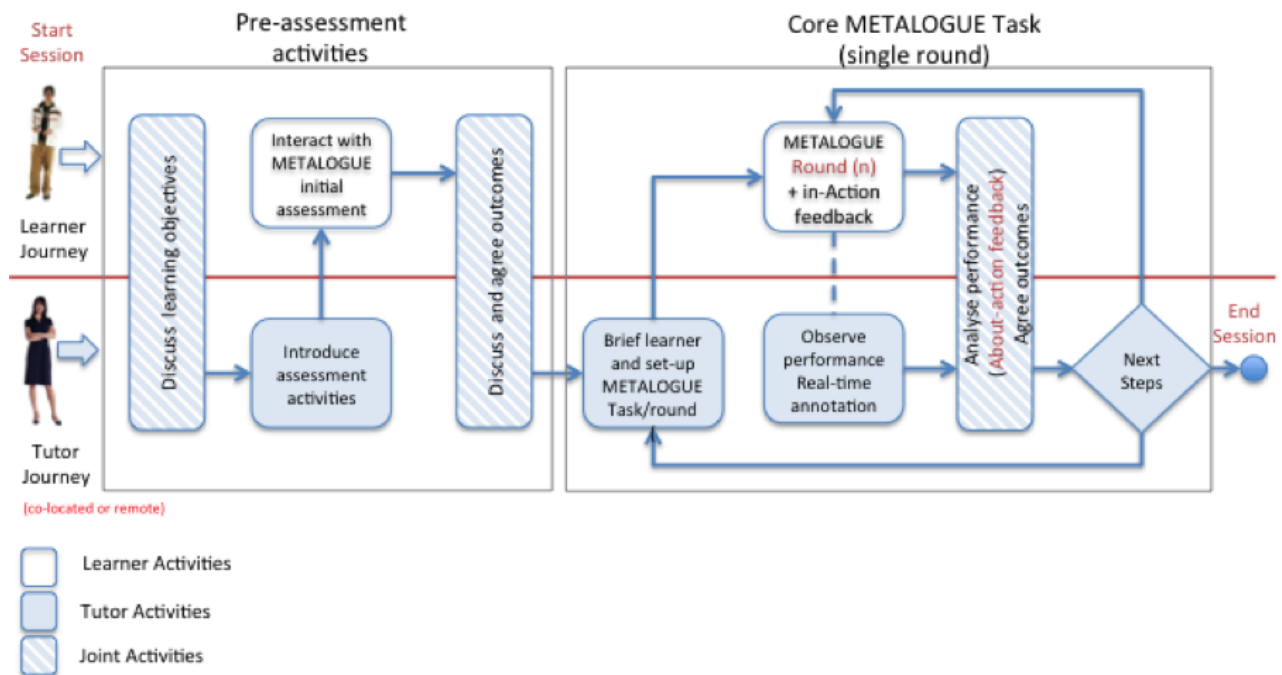


Figure 2. Learner and Tutor Journeys.

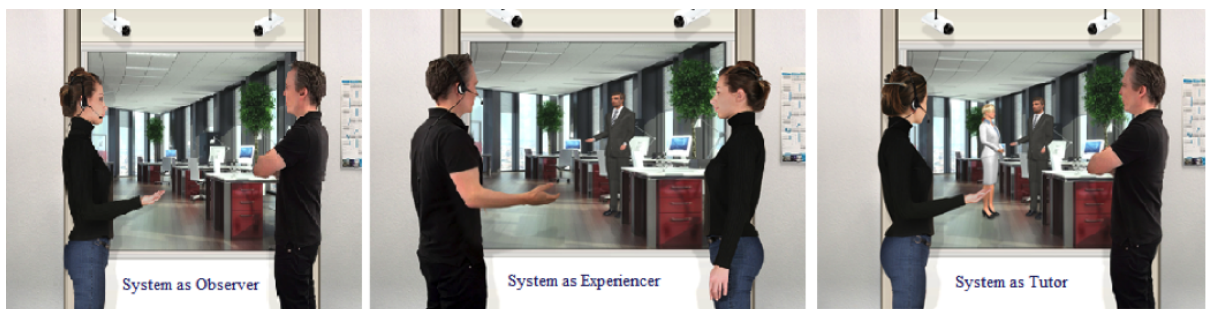


Figure 3. The Metalogue System as Observer, Experienter and Tutor.

behaviour as it happens, so as to optimize the immediately following action), and reflection-about-action (reflection after the event, to review, analyse, and evaluate the situation, so as to gain insight for improved practice in future).

The Metalogue feedback design consists of feedback given in-action (i.e. informing learners immediately how they are performing and how they can directly adapt to improve their performance) and about-action (i.e. informing learners about how they performed after finishing their task enabling them to adapt subsequent tasks). Consequently the Metalogue in-action feedback focuses on aspects of argument delivery. The Metalogue system can currently provide feedback on the use of certain hand and arm gestures, posture, volume, and tone of voice of the learner, and make suggestions to remedy inappropriate behaviour such slouching or speaking too softly. The system makes use of

several criteria to generate beliefs about the level of appropriateness of the learner's behaviour given certain interactive conditions (see Table 1). In other words, the system does not only understand what is happening in the dialogue as described in Section 3, but also interprets patterns of behaviour to infer to what extent the learner's actual behaviour deviates from the prescribed optimal behaviour.

The Metalogue system constantly generates a wide range of performance analysis, but real-time feedback is carefully balanced to avoid cognitive overload and/or disengagement whilst the learner is performing. For this reason, in-action feedback is currently envisioned in the form of a small number of simple signs and symbols displayed on-screen [43]. The goal of the given feedback is to help the learner become aware of their strengths and weaknesses, to enable a learner "to understand which aspects are of relevance and,

Feedback aspect		Debate	Negotiation
Presentation	speech volume	too loud (>60 decibel); too soft (<30 decibel)	
	hand and arm position posture	arms crossed; hands invisible (e.g. in the pockets, behind the back slouching)	
Interaction	turn management	abrupt and frequent interruption of other speaker (overlapping speech)	
	time management	long turns (>2 minutes)	long turns (>3 utterances)
Content	structure	missing/unmarked justifications/evidence (no discourse marker detected)	consecutive rejections (DeclineOffer sequences of length >1)
	semantics	irrelevant arguments (out-of-domain vocabulary)	non-cooperative negotiation moves (no BargainDown; no Concession acts)
	outcome	no agreement on any learner's argument (100% difference in participants' final debate states)	dead-lock situation; termination (Withdraw, Exit or BlockAgreement acts detected)

Table 1. Overview of the feedback on inappropriate behaviour as detected by the Metalogue system. Where possible, examples are provided in brackets.

ultimately, be able to recognise these aspects in their performance or the performance of others" [52]. While the in-action feedback is appropriately constrained, the about-action feedback provides the full range of possible feedback via a reflection dashboard that enables both tutor and learner to review and analyse the performance moment-by-moment. This is designed to build on the in-action feedback, taking the learner one step further. Thus the feedback given enables the learner to review, analyse, and evaluate the recent performance, and gain insight for improved practice in the future. The goal of this about-action feedback is to support the learner's self-monitoring, self-regulation and self-reflection.

In the "Experiencer" and "Tutor" roles (see Section 2), Metalogue gives in-action feedback on the content of the negotiation by evaluating the perceived cooperativeness of the actions of the learner and her trading partner. By providing this information as in-action feedback, the learner can directly observe how her actions affect her trading partner. In addition, the learner can develop the skills to distinguish between situations in which there is a small zone of possible agreements (ZOPA) [39], from situations in which she is negotiating with an uncooperative trading partner. The use of an avatar to represent the virtual agent allows us to administer this feedback in a unique way: by permitting the learner to see the contents of the agent's "mind." As the learner negotiates with the agent, the system displays feedback about how the agent is interpreting her actions through observable differences in her behaviour. This allows the agent to communicate to the learner - in real time - what it thinks the learner is doing and whether that action is considered to be cooperative or aggressive. In addition, the agent stores what she would have done herself instead to be reviewed as about-action feedback.

While in-action feedback necessarily only involves feedback that does not reveal the goals of the learner's trading partner, the learner is allowed to inspect both sides of the negotiation during about-action feedback. Currently, Metalogue provides about-action feedback on the content of the success of the negotiation for both negotiators in terms of how close the negotiated agreement is to reaching the highest possible score for each negotiator. In addition, the Metalogue system provides feedback on the cooperative success of the negotiation. For example, when negotiators reach an outcome that is not Pareto efficient, they "leave money on the table" [39] since there is an alternative outcome that increases the score of one negotiator without decreasing the score of her trading partner. In such a situation, the Metalogue system provides a concrete outcome that would have been at least as acceptable for the negotiation partner as the actual negotiated outcome, but would also have yielded the learner a higher score. This kind of feedback should allow learners to avoid similar situations through more (honest) communication about the goals of negotiators.

In addition, the Metalogue system stores each action that changes the state of the negotiation. The reflection dashboard allows both tutor and learner to replay the negotiation to analyse the learner's performance based on the learner's score, perceived cooperativeness, and trustworthiness in a moment-by-moment fashion. In future implementations of the Metalogue system, the reflection dashboard would also show what the virtual agent would recommend the learner to do at each point in time.

To summarise, given the set of debating skills discussed above, and in order for the system to provide feedback on the learner's performance, the Metalogue system records to the following aspects: (1) presentational aspects such as voice quality, speaking rate, and overall posture orientation; (2)

interactional aspects such as turn, time, contact, and own communication management; and (3) aspects related to argument structure. Furthermore Metalogue supports debate skills training with a consistent feedback loop, i.e. real-time in-action feedback to raise awareness of currently trained aspects/behaviours and about-action feedback to trigger reflection on the previous training sessions and prepare for the following training sessions.

5.1. Learning Analytics: About-action Feedback Dashboard Design

The term Learning Analytics generally refers to the large-scale measurement, collection and analysis of learner data across different systems within learning organisations [47]. Here we use the term to refer to the measurement, collection and analysis of Metalogue generated learner data relating to the individual learner, including comparisons with aggregated learner data from within the system. This type of reflection and analysis can support educational stakeholders in becoming “aware” of their actions and learning processes. Endsley [18, 19] described being “aware” as a three level process consisting of the perception of elements in the current situation, the comprehension of the current situations and the projection of a future status. These three steps are seen as a prerequisite for making decisions and effectively performing tasks. Once people are aware of their situation, they can reflect on their actions, choose to adapt their behaviour if necessary, and engage in a process of continuous learning [45].

Here we discuss our vision for the about-action reflection dashboard based on two user scenarios: (1) immediate post performance review by tutor and learner together; (2) tutor or learner individually reviewing at leisure.

There are a wide variety of software tools available to analyse and/or visualise existing data [26]. However, of particular interest in the context of the Metalogue project is the use of visualisation tools as demonstrated by the Flashmeeting project [46]. Although the application has been developed to support online meetings, the analysis tools provide a useful illustration of how a multimodal system data and analysis could be organised in the form of a reflection dashboard. For example, it is possible to replay the complete interaction, visualising the actual video replay as well as the broadcasting distribution over time (i.e. who spoke when and for how long), and more detailed information such as chat events, specific content annotations, and broadcasting events such as interruptions. It is also possible to view analyses such as broadcast dominance (i.e. the ratio of contribution by the different participants) in the form of a pie chart, and analysis of

the interaction content in the form of a key word cloud. The key criteria for the visualization of Metalogue data are:

- Occurrences of an event (e.g. voice volume (too high, too low etc), confident posture) on a timeline
- Aggregation of a single event (e.g. time used)
- Occurrences of a number of events in relation to each other in time.
- Integrated overview of various events.

In addition, parts of the dashboard may need to be layered, particularly when an integrated event is shown, and the learner or tutor should be able to zoom-in into the underlying aspects for further clarification.

An example of the type of about-action analysis the Metalogue system can provide the tutor and learner is shown as a screen mock-up in Figure 4. There are a wide range of options for selecting and viewing the whole or segments of the learner’s performance during a debate round. Video material appears in the top left window with standard video controls immediately beneath. Below (central) is the timeline adapted from the Flashmeeting dashboard as discussed above. This shows the utterances of both the learner and her opponent, plus the Metalogue feedback events against the timeline of the video. Clicking on an utterance block or an event symbol will display the corresponding video segment in the top left window. Similarly, clicking a particular point on the timeline will locate that point in the video.

The tabs along the top of the timeline window allow the user to view different types of event symbol; for example, Figure 4 (central) shows the voice tab has been selected. Accordingly, the symbols along the timeline represent all the Metalogue feedback events relating to voice performance. As the video plays, the top right window displays in detail any feedback events located on the timeline, this includes feedback given in-action, i.e. during performance, and all other events detected by Metalogue during performance but not displayed at the time to avoid overloading the learner. In Figure 4 the feedback is shown as a stopwatch symbol, providing positive feedback to the learner that her speech rate at this point was at an ideal level. It also provides clarification of what the symbol represents and provides links for further exploration.

The lower window is intended to provide various kinds of analysis depending on the tab selected on the timeline window above. In this case, with the voice tab selected (Figure 4, central), the available analysis options for voice are displayed along the top of the lower analysis window i.e. pause, emphasis, volume etc. The “Overall” tab is shown as selected (lower-right) and the window displays an analysis of the learner’s

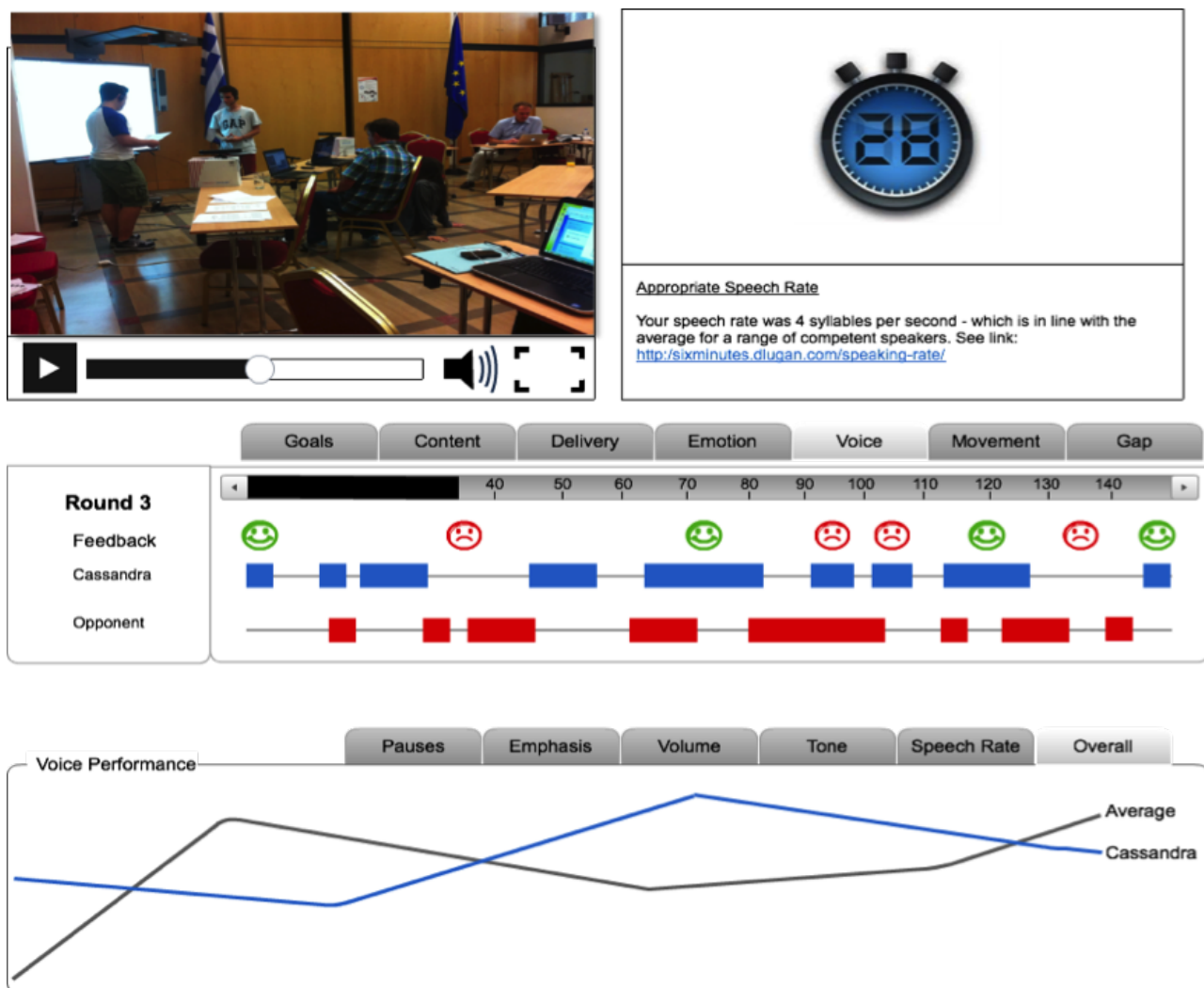


Figure 4. The Metalogue About-action dashboard screen mock-up showing video replay (top left), detected events on a timeline (centre) with related detailed advice (top right) and analysis (lower).

voice performance for the round against the average performance of other learners training with the same game parameters.

6. An immersive experience?

Debate and negotiation, whether human-human or human-virtual human, is by its very nature immersive; however, returning to the “digital immersion” criteria [15, 27, 41] mentioned in the introduction to this paper (i.e. situated learning, multiple perspectives, and transfer), we will review the proposed Metalogue functionality.

With regard to situated learning, Metalogue is being designed to support the 4C-ID pedagogic framework [51] which mandates attention to authentic whole tasks based on real life, organised in classes with variation and increasing complexity. The learner experience is also dynamic and engaging, with the system taking

the role of observer/audience and coach. In addition, it offers the option of a virtual debate partner able to employ natural language interaction and different styles of delivery (e.g. aggressive, conciliatory etc.).

Both in-action and about-action feedback offer the learner multiple alternative perspectives on their performance enabling them to become aware of certain behaviours and make choices about how to respond when necessary. For example, in-action: resetting body posture that has become inappropriate or gaze that has become averted; about-action: recognising and understanding debate strategies and how to employ them.

Finally, in terms of transferability, Metalogue is a mixed-reality system therefore the context of the learner is never entirely removed from the real world. Also the simulations involve realistic debating scenarios that allow the tutor to determine the topic and set

the extent of the challenge, thus learning outcomes are envisaged to be highly transferrable.

7. Vision into reality

The Metalogue project is being realised in three incremental prototype development cycles working towards the conceptual vision outlined in the earlier sections of this paper. Here we discuss the current status of Metalogue development i.e. what we have already achieved in terms of development, integration, deployment and user evaluation.

7.1. Prototype Development and Evaluation

In order to both trail the Metalogue concept with users and to gather corpus data, we deployed a Wizard of OZ version of Metalogue as a pre-pilot in the early months of the project. Learners enacted a debate scenario while being recorded by Kinect cameras and various microphones. At the same time, project members used a manual application to display appropriate real-time feedback in the form of red or green lights to help the participants improve their performance. A wide range of insight was gleaned from this experiment which has been fed in to the design of subsequent pilots and evaluation strategies.

Our first integrated prototype focused again on the debate scenario with Metalogue acting in the role of observer and coach. Referring to the User Journey illustrated in Figure 2, this involved three debate students from the Hellenic Youth Parliament in Athens performing a limited Core Metalogue Task; namely a debate-style interaction between a proponent and opponent on the subject of banning smoking in public areas. Learners completed the same round twice in different opponent/proponent pairings and each round was followed by a review with their tutor using the About-action feedback interface. Both tutors and learners provided informal feedback on their experience.

At this early stage, only a limited number of components were available for integration (see Figure 5 for details of the integrated components in the pilot 1 architecture). However, this time the system was able to provide automated In-action feedback in the form of a limited range of words and symbols relating to the learners' body stance and gestures. Also proprietary video annotation software was used as a provisional About-action feedback interface providing video replay of specific feedback events via a clickable time-line (see Section 5.1). As discussed earlier in Section 5, In-action feedback must be carefully balanced to avoid exceeding the cognitive capacity of the learner as they are performing, although Metalogue is designed to detect a much wider range behaviours, verbal and physical, potentially useful for the learner's development. To

illustrate this feature the about-action feedback time-line was set up to include not only the in-action events but additional incidents where the volume of the learner's voice dropped below normative levels.

Extensive testing of the integrated software and devices ensured the first prototype was stable and suitable for conducting a user evaluation. Despite the prototype's limitations, participants found the Metalogue concept (as illustrated by the pilot) to be acceptable and learners responded well to the In-action feedback provided. However, the provisional About-action interface was found to be difficult to navigate and interpret, highlighting the need to implement the planned bespoke interface (See Figure 4) in the subsequent pilots. Again a wide range of insight was gathered and used to shape the design of subsequent pilots.

Progress on the development and integration of components (see Figure 6 for the proposed architecture) has enabled us to design the forthcoming second pilot around the negotiation scenario, as discussed in Section 2.2, with Metalogue taking the role of a visible negotiation partner in the form of a human-size avatar. The learner, acting as a representative of the business community, will be able to negotiate the terms of a smoking ban with the avatar who represents the interests of a local authority. The parameters of the negotiation will allow for the trading of multiple options (i.e. smoking on public transport, scope of advertising campaign etc.) and In-action feedback will provide learner coaching across different modalities.

Alongside the prototyping of Metalogue in academic debate and negotiation learning contexts, we are exploring its potential to support corporate training, in particular, the coaching of call-centre agents in effective customer handling. We are also considering how the system might be ported into different language environments, specifically German and Greek.

7.2. Challenges

From the outset Metalogue was an ambitious vision and has therefore taken on a significant number of challenges. Here we discuss three of the key issues Metalogue researchers are currently working to address.

Challenge 1: Limitations of sensor tracking, recognition and natural language processing (NLP) technologies and error handling strategies. Despite significant advances in conversational technologies such as speech recognition, synthesis, natural language processing, and dialogue interaction management, the Metalogue system is still a long way from being a robust, fully automatic system that is able to understand debate arguments and negotiation moves accurately enough to achieve human-like performance in debates or negotiations.

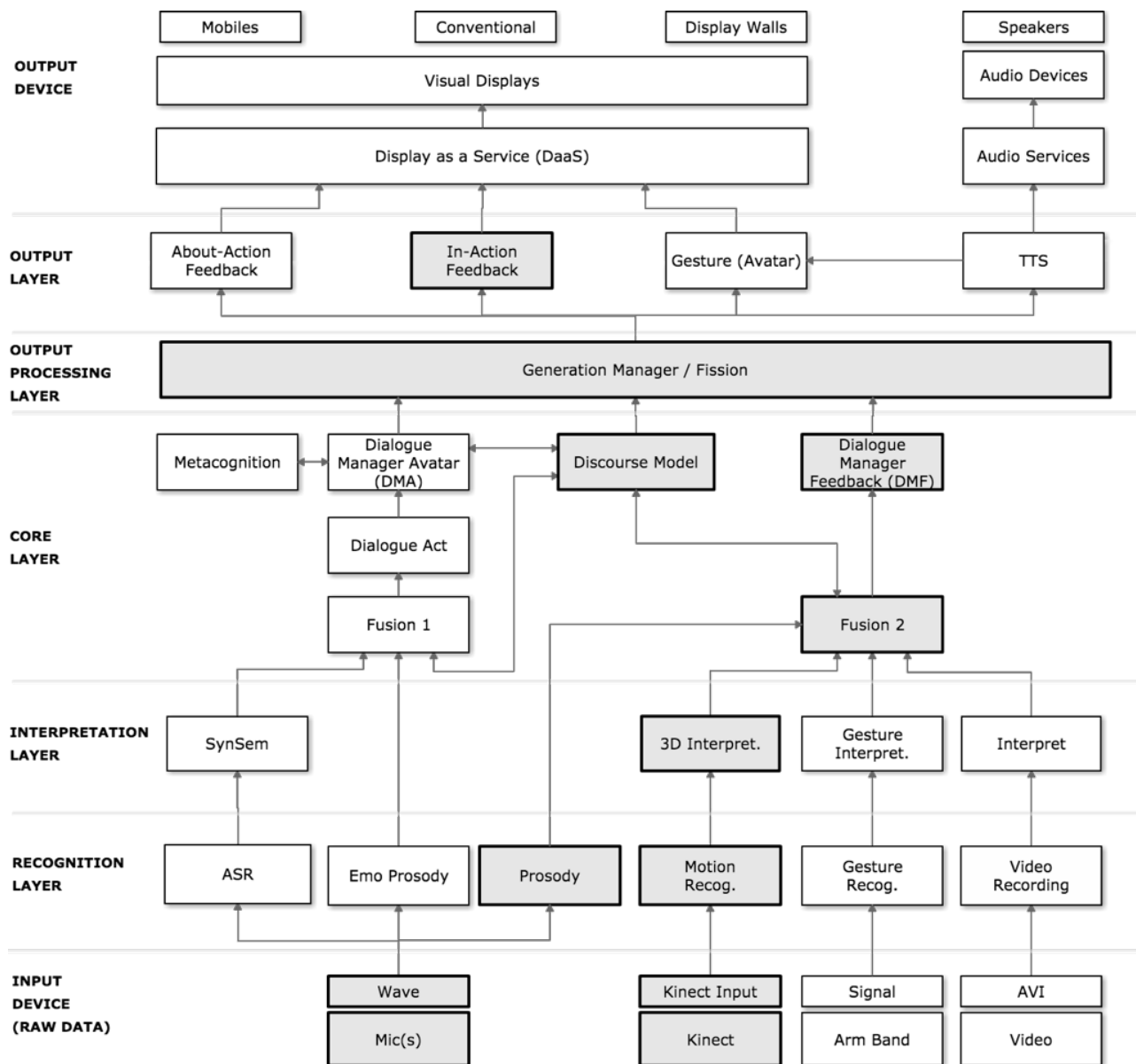


Figure 5. The Metalogue Pilot 1 Architecture: Shaded boxes indicate the components integrated to form the first prototype.

Each individual module of the dialogue system produces output that may contain errors of various types. For example, we currently observe some drop in performance in each of the following Metalogue modules: ASR has about a 28% Word Error Rate (WER); 7% of errors are detected in the syntactic and semantic parsers output; the recognition error rate of dialogue acts is approximately 14%; and only 25% of argument structures are recognized successfully, although the performance is better on explicitly marked argument relations. Each error type has a different impact on the overall system performance. There is growing evidence that many real-life applications using conversational technologies do not require 100% accuracy to be useful. Humans when communicating with each other

also do not show a perfect and deep understanding of each others contributions. Moreover, they may be deliberately vague in expressing their communicative intentions.

One particular challenge in the Metalogue system is that errors made by one module may propagate further through the system. As dialogue systems designers, Metalogue researchers are aware of typical processing mistakes and constantly analyse and assess each module's performance in an attempt to avoid such situations and, where the nature of such errors are known, reduce their negative impact in further processing steps. In this respect, complementary modalities play an important role in the Metalogue system. Complementary modalities not only improve the user experience, but also act as

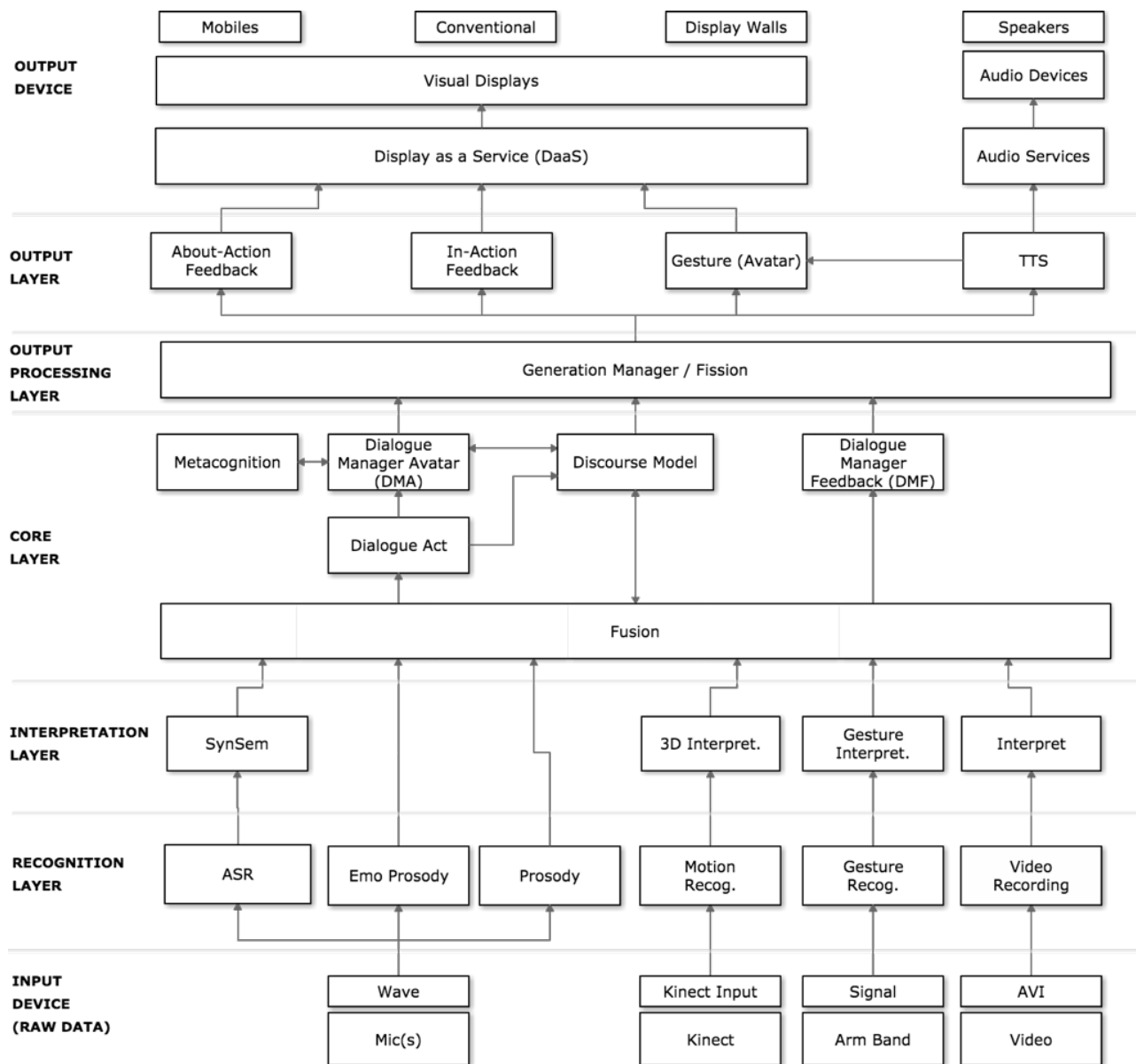


Figure 6. The Metalogue Pilot 2 proposed architecture.

additional information sources for the user's dialogue contributions. For example, visual body movements, facial expressions, and gaze direction may play a variety of functions in interactive situations [35].

As the last and most critical component, the dialogue manager is particularly vulnerable to errors that propagate through the Metalogue system. For this reason, the dialogue manager makes use of built-in error recovery strategies, ensuring that the system generates behaviour that is adequate and appropriate given the dialogue setting (debate or negotiation) and communication style (cooperative or non-cooperative). For example, if the system believes that the user has performed an action that it failed to recognize completely, a good strategy would be to ask the user

to repeat or rephrase what he just said. At a higher processing level, if the Metalogue system successfully recognized the user's action, but the interpreted action does not make any sense given the current system's state (e.g. irrelevant, ambiguous, or actions that conflict with previous actions), the system will initiate a relevant clarification sub-dialogue, see [34] for more details on these and other dialogue strategies and styles.

Challenge 2: Data quantity and quality. In Metalogue, debate and negotiation interactive multimodal data plays a crucial role since many modules deploy data-oriented approach for a systematic analysis and modelling of a variety of phenomena, such as negotiation

process, argumentation, feedback, and recurring structural patterns, for training statistical machine learning algorithms for semantic processing, for automatic speech and gesture recognition, prediction of dialogue acts. For instance, one of the challenges to create a computational agent within the Metalogue system that can engage in negotiation with a human learner, was to obtain data about the way people negotiate on an action-to-action basis. Typical negotiations among artificial agents make use of hard offers. This means that an agent that makes an offer is making a commitment to that offer. If the trading partner agrees to the offer, both agents are bound to the agreement. Humans, on the other hand, usually make tentative offers that do not have the same binding quality. In extreme cases, negotiators may even make offers that they are unwilling to accept themselves. To inform the cognitive agent in Metalogue, we recorded free-form negotiations between human participants to determine what kind of actions people take, and what levels of commitment human negotiators distinguish. Debate data has been collected in Wizard-of-Oz (WoZ) setting and in human-human scenario-based interactions simulating participants behaviour close to real life conditions. Modern sensors for tracking multimodal behaviour have been used as described in Section 3. The amount of data collected within the project so far are 8 hours for debate and 8 hours of negotiation dialogues. Obviously, quantity and the quality of some sessions (e.g. missing data due to the equipment and some tracking software limitations) is insufficient for many tasks. For indication, to train a robust automatic speech recognition system about 300 hours of data with vocabulary size of 40k words is needed to achieve the performance of WER below 25%. Moreover, in Metalogue, we had English none-native speakers as dialogue participants resulting in many accents variations posing a challenge for robust acoustic modelling. In order to develop a stable and robust system, we used other collected spoken dialogue/non-dialogue data (appr. 254 hours). We also have trained multiple models using different approaches and diverse training criteria, as well as several state-of-the-art model adaptation techniques. Given the importance of such data for the community and time invested in its collection, the Metalogue corpus will be released for the research community at the end of 2016.

Challenge 3: Pedagogy. In the “Experiencer” and “Tutor” roles, a learner can review the consequences of individual actions in a negotiation on the internal state of the computational agent. However, it is important to provide learners with the appropriate level of feedback. While a learner could pinpoint the precise sequence of actions that leads to negative effects in a negotiation process, the goal of the Metalogue system is for the

learner to appreciate the uncertainty of the negotiation process as well. The challenge for the Metalogue system is to provide feedback on an action-by-action basis, such that the learner can acquire appropriate negotiation behaviour, which is typically presented as general rules and guidelines [39].

A new evaluation method has been designed and tested to validate to what extent the system’s tutoring interventions correspond to that of human tutors. For this, we conducted a series of Wizard-of-Oz (WoZ) experiments whose main goal was to study the effects of tutoring interventions. The output from this experiments has been used as simulated input for our tutoring system. Two human Wizard tutors provided feedback on the presentational aspects of a trainee’s performance in real time by pressing a red button for negative feedback, e.g. “talks too loud”, “wrong posture” etc., and a green button for positive feedback. Two debate sessions were evaluated with the total duration of 22 minutes consisting of 57 turns of 4 different speakers, and comprising 426 utterances. Time-stamped human Wizard’s and automatically generated tutoring interventions events were logged and compared. We observed that human and system interventions differ both quantitatively and qualitatively. The system generated about 50% more feedback messages, with a significantly higher portion of negative feedback than human tutors do. There are also differences between tutors’ judgements with an agreement between them reaching less than 30%. It should be noted here, that human tutors involved in this preliminary study were of different backgrounds, e.g. psychologists and pedagogical experts, but not professional debate or negotiation trainers. This small-scale experiment however provided useful insights into evaluation procedure and criteria. For instance, the study indicated that the amount, type and complexity of feedback which is appreciated most and considered useful is a still largely open concern. Thus, user-based evaluation and usability testing should be performed on a large scale involving both trainees and tutors. From the evaluation with trainees insights can be gained on what skills and what aspects are most important for them to master, and from the evaluation with tutors what type, amount and timing of interventions they provide lead to the best learning outcome.

7.3. In Conclusion

The vision elaborated above is a challenging, multi-disciplinary endeavour and a work in progress. However, the outcomes have the potential to advance our understanding of metacognitive training and learning analytics on one level, and conversational human-machine interaction, dialogue interfaces on

another - perhaps moving us one small step towards the visionary capabilities embodied in Jonze' Her [24].

Acknowledgements

The authors would like to thank all Metalogue staff who contributed in word and writing and in many discussions. The underlying research project is partly funded by the Metalogue project; a Seventh Framework Programme collaboration funded by the European Commission, grant agreement number: 611073 <http://www.metalogue.eu>.

7.4. Bibliography

References

- [1] ALEVEN, V., MCLAREN, B., ROLL, I. and KOEDINGER, K. (2006) Toward meta-cognitive tutoring: A model of help-seeking with a cognitive tutor. *International Journal of Artificial Intelligence in Education* **16**: 101–128.
- [2] AMANOVA, D., PETUKHOVA, V. and KLAOW, D. (2016) Creating annotated dialogue resources: Cross-domain dialogue act classification. In *Proceedings 9th International Conference on Language Resources and Evaluation (LREC 2016)* (ELRA, Paris).
- [3] ARGYLE, M. (1994) *The psychology of interpersonal behaviour* (Penguin Books, London).
- [4] AZEVEDO, R., WITHERSPOON, A., CHAUNCEY, A., BURKETT, C. and FIKE, A. (2002) MetaTutor: A metacognitive tool for enhancing self-regulated learning. In PIRRONE, R., AZEVEDO, R. and BISWAS, G. [eds.] *Cognitive and Metacognitive Educational Systems: Papers from the AAAI Fall Symposium (FS-09-02)*.
- [5] BAKER, R.S.J., CORBETT, A.T., KOEDINGER, K.R. and ROLL, I. (2006) Generalizing detection of gaming the system across a tutoring curriculum. In *Intelligent Tutoring Systems: 8th International Conference, ITS 2006, Jhongli, Taiwan, June 26-30, 2006* (Springer), *Lecture Notes in Computer Science* **4053**: 402–411.
- [6] BOSER, B.E., GUYON, I.M. and VAPNIK, V.N. (1992) A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory* (ACM): 144–152.
- [7] BRANSFORD, J.D., BROWN, A.L. and COCKING, R.R. (2000) *How People Learn: Brain, Mind, Experience, and School* (National Academy Press).
- [8] BROWN, A.L. and CAMPIONE, J.C. (1996) Psychological theory and the design of innovative learning environments: On procedures, principles, and systems. In SCHAUBLE, L. and GLASER, R. [eds.] *Innovations in Learning: New Environments for Education* (Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc), 289–325.
- [9] BROWN, A.L., PALINCAS, A.S. and ARMBRUSTER, B.B. (1984) Instructing comprehension-fostering activities in interactive learning situations. *Learning and comprehension of text* : 255–286.
- [10] BUNT, A. and CONATI, C. (2003) Probabilistic student modelling to improve exploratory behaviour. *User Modeling and User-Adapted Interaction* **13**(3): 269–309. doi:10.1023/A:1024733008280.
- [11] CHI, M.T., SILER, S.A., JEONG, H., YAMAUCHI, T. and HAUSMANN, R.G. (2001) Learning from human tutoring. *Cognitive Science* **25**(4): 471–533. doi:10.1016/S0364-0213(01)00044-1.
- [12] CODE, J.R. (2010) *Assessing Agency for Learning: Agency for Learning and Measuring Agency for Learning: Validating the Agency for Learning Questionnaire (AFLQ) and Agency as a Mediator of Academic Achievement*. Ph.D. thesis, Simon Fraser University. <http://summit.sfu.ca/item/11308>.
- [13] DE WEERD, H., BROERS, E. and VERBRUGGE, R. (2015) Savvy software agents can encourage the use of second-order theory of mind by negotiators. In *Proceedings of the 37th Annual Conference of the Cognitive Science Society*. (Pasadena, California): 542–547.
- [14] DE WEERD, H., VERBRUGGE, R. and VERHEIJ, B. (2013) Higher-order theory of mind in negotiations under incomplete information. In *PRIMA 2013: Principles and Practice of Multi-Agent Systems*: 101–16.
- [15] DEDE, C. (2009) Immersive interfaces for engagement and learning. *Science* **323**(5910): 66–69. doi:10.1126/science.1167311.
- [16] DEDE, C. (2012) Customization in immersive learning environments. In DEDE, C. and RICHARDS, J. [eds.] *Digital Teaching Platforms: Customizing Classroom Learning for Each Student* (New York, United States: Teachers College Press), 282–297.
- [17] DUNCAN, S. (1970) Towards a grammar for floor apportionment: A system approach to face-to-face interaction. In *Proceedings of the Second Annual Environmental Design Research Association Conference*: 225–236.
- [18] ENDSLEY, M.R. (1995) Measurement of situation awareness in dynamic systems. *Human Factors* **37**(1): 65–84. doi:10.1518/001872095779049499.
- [19] ENDSLEY, M.R. (2000) Theoretical underpinnings of situation awareness: A critical review. In ENDSLEY, M.R. and GARLAND, D.J. [eds.] *Situation Awareness Analysis and Measurement* (Boca Raton, FL, United States: CRC Press), 3–32.
- [20] GAMA, C. (2004) Metacognition in interactive learning environments: The reflection assistant model. In LESTER, J.C., VICARIO, R.M. and PARAGUACU, F. [eds.] *Intelligent Tutoring Systems: 7th International Conference, ITS 2004, Maceio, Alagoas, Brazil, August 30 - September 3, 2004* (Springer), *Lecture Notes in Computer Science* **3220**, 668–677.
- [21] HATTIE, J. (2008) *Visible learning: A synthesis of over 800 meta-analyses relating to achievement* (London, England: Routledge).
- [22] HEYLEN, D. (2006) Head gestures, gaze and the principles of conversational structure. *International journal of Humanoid Robotics* **3**(3): 241–267.
- [23] JOHNSON, W.L. and VALENTE, A. (2009) Tactical language and culture training systems: Using AI to teach foreign languages and cultures. *AI Magazine* **30**(2): 72–83. doi:10.1609/aimag.v30i2.2240.
- [24] JONZE, S. (2014) Her. Elite Films, Zurich.
- [25] KENDON, A. (2004) *Gesture: visible action as utterance* (Cambridge University Press).
- [26] KRAAN, W. and SHERLOCK, D. (2013) Analytics tools and infrastructure. *CETIS Analytics Series* **1**(11).

- [27] LESSITER, J., FREEMAN, J., KEOGH, E. and DAVIDOFF, J. (2001) A cross-media presence questionnaire: The ITC-sense of presence inventory. *Presence* 10(3): 282–297. doi:10.1162/105474601300343612.
- [28] LIN, R., GAL, Y., KRAUS, S. and MAZLIAH, Y. (2014) Training with automated agents improves people's behavior in negotiation and coordination tasks. *Decision Support Systems* 60(1): 1–9. doi:10.1016/j.dss.2013.05.015, URL <http://dx.doi.org/10.1016/j.dss.2013.05.015>.
- [29] LYBBERT, B. (1985) What should be the goals of high school debate? Paper presented at the National Forensic League Conference on the State of Debate, Kansas City, MO.
- [30] MALCHANAU, A., PETUKHOVA, V., BUNT, H. and KLAOW, D. (2015) Multidimensional dialogue management for tutoring systems. In *Proceedings of the 7th Language and Technology Conference (LTC 2015)* (Poznan, Poland).
- [31] MARZANO, R.J. (1998) A theory-based meta-analysis of research on instruction. (Aurora, Colorado: Midcontinent Regional Educational Laboratory).
- [32] NYE, B.D., GRAESSER, A.C. and HU, X. (2014) AutoTutor and family: A review of 17 years of natural language tutoring. *International Journal of Artificial Intelligence in Education* 24(4): 427–469.
- [33] PELDSZUS, A. and STEDE, M. (2013) From argument diagrams to argumentation mining in texts: A survey. *International Journal of Cognitive Informatics and Natural Intelligence* 7(1): 1–31.
- [34] PETUKHOVA, V. (2011) *Multidimensional Dialogue Modelling*. PhD dissertation (The Netherlands: Tilburg University).
- [35] PETUKHOVA, V. and BUNT, H. (2010) Towards an integrated scheme for semantic annotation of multimodal dialogue data. In *Proceedings of the seventh international conference on language resources and evaluation* (Paris: ELRA): 2556–2563.
- [36] PETUKHOVA, V., MALCHANAU, A. and BUNT, H. (2015) Modelling argumentative behaviour in parliamentary debates. In *Proceedings of the 15th Workshop on Computational Models of Natural Argument, Principles and Practice of Multi-Agent Systems Conference (PRIMA 2015)* (Bertinoro, Italy).
- [37] PETUKHOVA, V., STEVENS, C., DE WEERD, H., TAATGEN, N., CNOSSEN, F. and MALCHANAU, A. (2016) Modelling multi-issue bargaining dialogues: data collection, annotation design and corpus. In *Proceedings 9th International Conference on Language Resources and Evaluation (LREC 2016)* (ELRA, Paris).
- [38] PFLEGER, N. (2007) *Context-based Multimodal Interpretation: An Integrated Approach to Multimodal Fusion and Discourse Processing*. Ph.D. thesis, Universität des Saarlandes.
- [39] RAIFFA, H., RICHARDSON, J. and METCALFE, D. (2002) *Negotiation Analysis: The Science and Art of Collaborative Decision Making* (Cambridge, MA: Belknap Press).
- [40] ROLL, I., ALEVEN, V., McLAREN, B.M. and KOEDINGER, K.R. (2007) Can help seeking be tutored? searching for the secret sauce of metacognitive tutoring. In LUCKIN, R., KOEDINGER, K. and GREER, J. [eds.] *Artificial Intelligence in Education: Building Technology Rich Learning Contexts that Work* (IOS Press), *Frontiers in Artificial Intelligence and Applications* 158: 203–210.
- [41] SADOWSKI, W. and STANNEY, K. (2002) Presence in virtual environments. In HALE, K.S. and STANNEY, K.M. [eds.] *Handbook of Virtual Environments: Design, Implementation, and Applications* (Lawrence Erlbaum Associates Publishers), 791–806.
- [42] SCHEFLEN, A. (1964) The significance of posture in communication systems. *Psychiatry* 17: 316–331.
- [43] SCHNEIDER, J., BÖRNER, D., VAN ROSMALEN, P. and SPECHT, M. (2014) Presentation trainer: A toolkit for learning non-verbal public speaking skills. In RENSING, C., DE FREITAS, S., LEY, T. and MUNOZ-MERINO, P. [eds.] *Open Learning and Teaching in Educational Communities: 9th European Conference on Technology Enhanced Learning, EC-TEL 2014, Graz, Austria, September 16-19, 2014, Proceedings* (Springer), *Lecture Notes in Computer Science* 8719, 522–525.
- [44] SCHNEIDER, J., BÖRNER, D., VAN ROSMALEN, P. and SPECHT, M. (2015) Augmenting the senses: A review on sensor-based learning support. *Sensors* 15(2): 4097–4133. doi:10.3390/s150204097.
- [45] SCHÖN, D.A. (1983) *The Reflective Practitioner: How Professionals Think in Action*, 5126 (Basic Books, NY).
- [46] SCOTT, P., TOMADAKI, E. and QUICK, K. (2007) The shape of online meetings. *The International Journal of Technology, Knowledge, and Society* 3(4): 1–16.
- [47] SIEMENS, G. (2012) Learning analytics: envisioning a research discipline and a domain of practice. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (ACM): 4–8.
- [48] STEVENS, C.A., TAATGEN, N. and CNOSSEN, F. (2016) Instance-based models of metacognition in the prisoner's dilemma. *Topics in Cognitive Science* 8(1): 322–334.
- [49] TEUFEL, S. (1999) *Argumentative Zoning: Information Extraction from Scientific Text*. Ph.D. thesis, University of Edinburgh.
- [50] TUMPOSKY, N.R. (2004) The debate debate. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas* 78(2): 52–56.
- [51] VAN MERRIËNBOER, J.J. and KESTER, L. (2008) Whole-task models in education. In MERRILL, M.D., VAN MERRIËNBOER, J.J., DRISCOLL, M.P. and SPECTOR, J.M. [eds.] *Handbook of Research on Educational Communications and Technology* (Routledge/Taylor & Francis Group New York), 3, 441–456.
- [52] VAN ROSMALEN, P., BÖRNER, D., SCHNEIDER, J., PETUKHOVA, V. and VAN HELVERT, J. (2015) Feedback design in multimodal dialogue systems. *CSEDU* : 209–217.
- [53] VAPNIK, V. (2013) *The nature of statistical learning theory* (Springer Science & Business Media).
- [54] WAHLSTER, W. [ed.] (2006) *SmartKom: Foundations of Multimodal Dialogue Systems* (Berlin, Heidelberg: Springer).
- [55] WHITE, B.Y. and FREDERIKSEN, J.R. (1998) Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and instruction* 16(1): 3–118.
- [56] YANG, H.Y., ZHANG, H., XU, W., ZHANG, P.J. and XU, L.M. (2014) The application of kinect motion sensing technology in game-oriented study. *International Journal of Emerging Technologies in Learning (iJET)* 9(2): 59–63.

- [57] YU, H.F, HUANG, F.L. and LIN, C.J. (2011) Dual coordinate descent methods for logistic regression and maximum entropy models. *Machine Learning* **85**(1-2): 41–75.
- [58] ZHU, J., ZOU, H., ROSSET, S. and HASTIE, T. (2009) Multi-class adaboost. *Statistics and its Interface* **2**(3): 349–360.